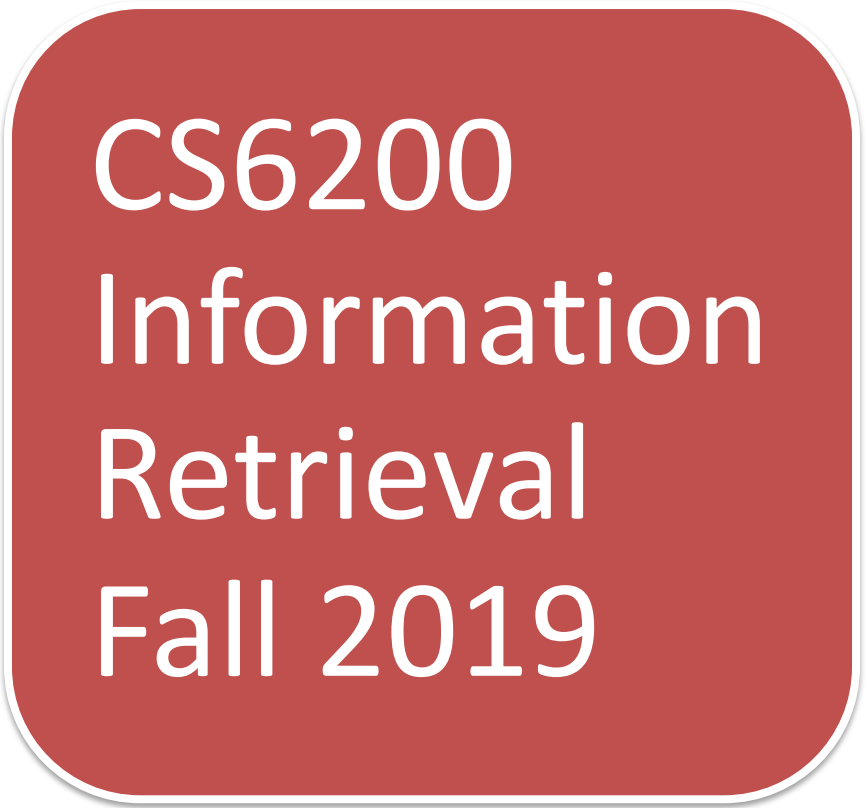


Northeastern University - Seattle

Khoury College of Computer Sciences

**Lecture 0: Course
Details & Expectations
Sep 9, 2019**

A red rounded rectangle with a white border and a subtle drop shadow, containing the course title in white text.

**CS6200
Information
Retrieval
Fall 2019**

CS 6200 02 Information Retrieval

Fall 2019

Instructor: Raman “Chandra” Chandrasekar

TA: Kartik Dave

- Class meetings: Mondays, 6:15pm – 9:15pm,
Sep 9th through Dec 9th
225 Terry Ave N/307
- Chandra Office Hours: 5pm – 6pm Mondays,
225 Terry Ave N/111

Search !

We'll explore and understand search, and hopefully
have fun doing that!

Chandra

- PhD in AI (Machine Translation), Tata Institute of Fundamental Research (TIFR), Mumbai, India.
- Worked at TIFR and at the National Centre for Software Technology. Helped create a group working on AI/Knowledge Based Systems, focused on natural language processing (NLP), intelligent tutoring and heuristic scheduling
- 1995 to 1998: At the Institute for Research in Cognitive Science, at the University of Pennsylvania. Worked on automatic sentence simplification and using NLP to improve search. Also at AT&T Labs Research.
- 1998 to 2010: Researcher at Microsoft (**msnSearch/Bing**) and Microsoft Research in Redmond. Worked on search, text mining, news dissemination, human computation etc.
- 2010 to 2015: At news dissemination startup Evri.com. Later, managed development of web-scale academic search products at ProQuest in Seattle.
- Now: Clinical Professor here; also, consultant. Working on text classification, search, and question-answering.



Kartik Dave

I'm a Computer Science Masters student having interest in the field of Big Data, Distributed Systems and Information Retrieval. I started with my master's from the Boston campus and later moved to Seattle. Also, while working on my master's I interned with Worldpay and AWS.

I love to travel and hike. My favorite experience of hiking is Precipice Trail, Acadia National Park.

I look forward to working with all of you this semester and having all the fun technical discussions.

Office Hours: TBA

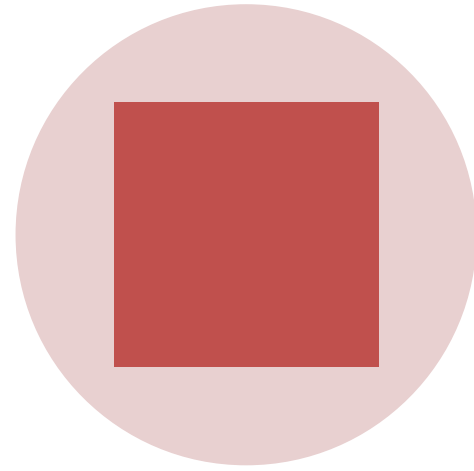
CS6200 Fall 2019



Introductions: You



YOUR NAME



SOMETHING ABOUT
YOURSELF

Fill in Survey

My Goals



Not to lecture “at” you, but to facilitate learning and encourage you to think and apply what you learn



Want to make sure concepts get across



Focus on web search, with an emphasis on what ‘real’ search engines do.

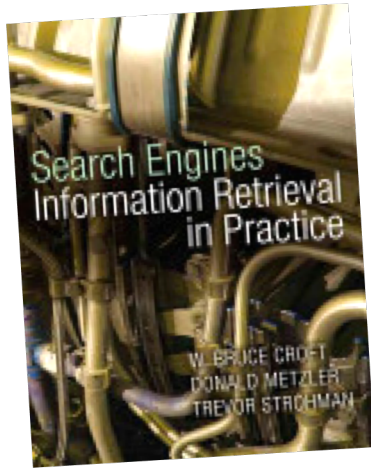


Questions always welcome during class.



I’ll also stop periodically for questions.

Texts



Search Engines: Information Retrieval in Practice. Bruce Croft, Donald Metzler and Trevor Strohman, Addison-Wesley. 2010. ISBN: 978-0-136-07224-9 [CMS]

Readings assigned from this.

Others:

- Introduction to Information Retrieval, Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Cambridge University Press. 2008. ISBN: 0521865719. [MRS]
- Information Retrieval: Implementing and Evaluating Search Engines, Stefan Büttcher, Charles L.A. Clarke, and Gordon V. Cormack, MIT Press. 2010. ISBN: 978-0-262-02651-2

Piazza



Piazza will be used for Lecture slides, Assignments, Discussions



Piazza signup link: <https://piazza.com/northeastern/fall2019/cs6200seattle>

Access Code: **SeattleIR**



Piazza discussion board:
<https://piazza.com/northeastern/fall2019/cs6200seattle/home>

Grading



1 Final Exam (20% of course grade)



5 homework assignments. (70% of course grade). More on next slide.



Quiz (almost) every week. (Max 10% of course grade). Several questions with short or multiple-choice answers. Hopefully, a fun, learning experience. *Gives me feedback on learning.* Each quiz worth 1%, best 10 contribute to grade.

Homework Assignments



Four assignments (2 weeks, 15% each):

implementation of a search engine component
Outputs: Code, output from specific data, note on implementation and results.



One assignment (10%):

Will involve a study, report and presentation



Want you to learn by doing, working through solutions on your own.



OK to discuss the problem with me, TA, classmates, but work on your own code, solutions & write-up without reading or copying others' solutions.

Late Policy, Academic Integrity, Accommodations



Assignments due as indicated at 9am on due date.



3 slip days available to you. Otherwise late assignments penalized 10% for each calendar day of delay.



For other details about this, about Academic Integrity, and Accommodations for Students with Disabilities, please look at the syllabus and the Student Handbook.

Syllabus & Schedule

Week #	Date	Topic	Readings	Deadlines *
1	9/9/2019	1: Overview of IR 2: Search Engine Architecture	CMS Chapters 1 & 2, Brin & Page	
2	9/16/2019	3: Acquiring Data	CMS Chapter 3	Assignment1 released
3	9/23/2019	4: Transforming Data	CMS Chapter 4	
4	9/30/2019	5: Indexing & Ranking	CMS Chapter 5, Sections 5.1 – 5.6	Assignment1 due
5	10/ 7/2019	6: Query Processing	CMS Sections 5.7, & 4.5; Adv Operators	Assignment2 released
6	10/14/2019	7: Query Refinement	CMS Chapter 6, Sections 6.1, 6.2	
7	10/21/2019	8: Retrieval Models – 1	CMS Chapter 7, Sections 7.1, 7.2	Assignment2 due Assignment3 released
8	10/28/2019	9: Retrieval Models – 2	CMS Chapter 7, Sections 7.3, 7.5 --7.7	
9	11/ 4/2019	10: Evaluating Search Engines	CMS Chapter 8, except for Section 8.6	Assignment3 due Assignment4 released
10	11/11/2019	Lec. 10 continued, plus 11: User Experience	CMS Chapter 6, Sections 6.3, 6.4; Dziadosz & Chandrasekar	
11	11/18/2019	Lec. 11 continued	CMS Chapter 6, Sections 6.3, 6.4; Dziadosz & Chandrasekar	Assignment4 due Assignment5 released
12	11/25/2019	Guest Lecture/ 12: Specialized Search	CMS Chapter 10 and Chapter 11 – Sections 11.4, 11.5, 11.6	
13	12/ 2/2019	Student Presentations		Assignment5 due
14	12/9/2019	12: Specialized Search cont'd/ Final Exam		

* Quizzes not shown

Questions?